

Fusion of Sparse Reconstruction Algorithms for Multiple Measurement Vectors

Deepa K G, Sooraj K. Ambat, and K.V.S. Hari

Abstract

We consider the recovery of sparse signals that share a common support from multiple measurement vectors. The performance of several algorithms developed for this task depends on parameters like dimension of the sparse signal, dimension of measurement vector, sparsity level, measurement noise. We propose a fusion framework, where several multiple measurement vector reconstruction algorithms participate and the final signal estimate is obtained by combining the signal estimates of the participating algorithms. We present the conditions for achieving a better reconstruction performance than the participating algorithms. Numerical simulations demonstrate that the proposed fusion algorithm often performs better than the participating algorithms.

Index Terms

Compressed Sensing, Fusion, Sparse Signal Reconstruction, Multiple Measurement Vectors

I. INTRODUCTION

Consider the standard Compressed Sensing (CS) measurement setup where a K -sparse signal $\mathbf{x} \in \mathbb{R}^{N \times 1}$ is acquired through M linear measurements via

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{w}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ denotes the measurement matrix, $\mathbf{b} \in \mathbb{R}^{M \times 1}$ represents the measurement vector, and $\mathbf{w} \in \mathbb{R}^{M \times 1}$ denotes the additive measurement noise present in the system. The reconstruction problem, estimating \mathbf{x} from (1) using \mathbf{A} and \mathbf{b} , is known as Single Measurement

Deepa K G, Sooraj K. Ambat, and K.V.S. Hari are with Statistical Signal Processing Lab, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, 560012, India. (email: {deepa,sooraj,hari}@ece.iisc.ernet.in)

Vector (SMV) problem. In this work, we consider the Multiple Measurement Vector (MMV) problem [1] where we have L measurements: $\mathbf{b}^{(1)} = \mathbf{A}\mathbf{x}^{(1)} + \mathbf{w}^{(1)}$, $\mathbf{b}^{(2)} = \mathbf{A}\mathbf{x}^{(2)} + \mathbf{w}^{(2)}$, \dots , $\mathbf{b}^{(L)} = \mathbf{A}\mathbf{x}^{(L)} + \mathbf{w}^{(L)}$. The vectors $\{\mathbf{x}^{(l)}\}_{l=1}^L$ are assumed to have a common sparse support set. The problem is to estimate $\mathbf{x}^{(l)}$ ($l = 1, 2, \dots, L$). Instead of recovering the L signals individually, the attempt in the MMV problem is to simultaneously recover all the L signals. MMV problem arises in many applications such as the neuromagnetic inverse problem in Magnetoencephalography (a modality for imaging the brain) [2], [3], array processing [4], non-parametric spectrum analysis of time series [5], and equalization of sparse communication channels [6].

Recently many algorithms have been proposed to recover signal vectors with a common sparse support. Some among them are algorithms based on diversity minimization methods like $\ell_{2,1}$ minimization [7], and M-FOCUSS [1], greedy methods like M-OMP and M-ORMP [1], and Bayesian methods like MSBL [8] and T-MSBL [9].

However it has been observed that the performance of many algorithms depends on many parameters like the dimension of the measurement vector, the sparsity level, the statistical distribution of the non-zero elements of the signal, the measurement noise power etc. [9]. Thus it becomes difficult to choose the best sparse reconstruction algorithm without *a priori* knowledge about these parameters.

Suppose we have the sparse signal estimates given by various algorithms. It may be possible to merge these estimates to form a more accurate estimate of the original. This idea of fusion of multiple estimators has been proposed in the context of signal denoising in [10] where fusion was performed by plain averaging. Recently, Ambat *et al.* [11]–[15] proposed fusion of the estimates of sparse reconstruction algorithms to improve the sparse signal reconstruction performance of SMV problem.

In this paper, we propose a framework which uses several MMV reconstruction algorithms and combines their sparse signal support estimates to determine the final signal estimate. We refer to this scheme as *MMV-Fusion of Algorithms for Compressed Sensing* (MMV-FACS). We present an upper bound on the reconstruction error by MMV-FACS. We also present a sufficient condition for achieving a better reconstruction performance than any participating algorithm. By Monte-Carlo simulations we show that fusion of viable algorithms leads to improved reconstruction performance for the MMV problem.

Notations:

Matrices and vectors are denoted by bold upper case and bold lower case letters respectively. Sets are represented by upper case Greek alphabets and calligraphic letters. $\mathbf{A}_{\mathcal{T}}$ denotes the column sub-matrix of \mathbf{A} where the indices of the columns are the elements of the set \mathcal{T} . $\mathbf{X}_{\mathcal{T},:}$ denotes the sub-matrix formed by those rows of \mathbf{X} whose indices are listed in the set \mathcal{T} . \mathbf{X}^K is the matrix obtained from \mathbf{X} by keeping its K rows with the largest ℓ_2 -norm and by setting all other rows to zero, breaking ties lexicographically. $\text{supp}(\mathbf{X})$ denotes the set of indices of non-zero rows of \mathbf{X} . For a matrix \mathbf{X} , $\mathbf{x}^{(l)}$ denotes the l^{th} column vector of \mathbf{X} . $\hat{\mathbf{X}}_i$ denotes the reconstructed matrix by the i^{th} participating algorithm. The complement of the set \mathcal{T} with respect to the set $\{1, 2, \dots, N\}$ is denoted by \mathcal{T}^c . For two sets \mathcal{T}_1 and \mathcal{T}_2 , $\mathcal{T}_1 \setminus \mathcal{T}_2 = \mathcal{T}_1 \cap \mathcal{T}_2^c$ denotes the set difference. $|\mathcal{T}|$ denotes the cardinality of set \mathcal{T} . \mathbf{A}^\dagger and \mathbf{A}^T denote the pseudo-inverse and transpose of matrix \mathbf{A} , respectively. The (p, q) mixed norm of the matrix \mathbf{X} is defined as

$$\|\mathbf{X}\|_{(p,q)} = \left(\sum_i \|\mathbf{X}_{i,:}\|_p^q \right)^{1/q}$$

The Frobenius norm of matrix \mathbf{A} is denoted as $\|\mathbf{A}\|_F$.

II. PROBLEM FORMULATION

The MMV problem involves solving the following L under-determined systems of linear equations

$$\mathbf{b}^{(l)} = \mathbf{A}\mathbf{x}^{(l)} + \mathbf{w}^{(l)}, \quad l = 1, 2, 3, \dots, L \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ ($M \ll N$) represents the measurement matrix, $\mathbf{b}^{(l)} \in \mathbb{R}^{M \times 1}$ represents the l^{th} measurement vector, and $\mathbf{x}^{(l)} \in \mathbb{R}^{N \times 1}$ denotes the corresponding K -sparse source vector. That is, $|\text{supp}(\mathbf{x}^{(l)})| \leq K$ and $\mathbf{x}^{(l)}$ share a common support-set for $l = 1, 2, \dots, L$. $\mathbf{w}^{(l)} \in \mathbb{R}^{M \times 1}$ represents the additive measurement noise. We can rewrite (2) as

$$\mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W} \quad (3)$$

where $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(L)}]$, $\mathbf{W} = [\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots, \mathbf{w}^{(L)}]$, and $\mathbf{B} = [\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(L)}]$. For a matrix \mathbf{X} , we define

$$\text{supp}(\mathbf{X}) = \bigcup_{i=1}^L \text{supp}(\mathbf{x}^{(i)}).$$

In (3), we assume that \mathbf{X} is jointly K -sparse. That is, $|\text{supp}(\mathbf{X})| \leq K$. There are at most K rows in \mathbf{X} that contain non-zero elements. We assume that $K < M$ and K is known *a priori*.

III. FUSION OF ALGORITHMS FOR MULTIPLE MEASUREMENT VECTOR PROBLEM

In this paper, we propose to employ multiple sparse reconstructions algorithms independently for estimating \mathbf{X} from (3) and fuse the resultant estimates to yield a better sparse signal estimate. Let $P \geq 2$ denote the number of different participating algorithms employed to estimate the sparse signal. Let $\hat{\mathcal{T}}_i$ denote the support-set estimated by the i^{th} participating algorithm and let \mathcal{T} denote the true-support-set. Denote the union of the estimated support-sets as Γ , i.e., $\Gamma \triangleq \cup_{i=1}^P \hat{\mathcal{T}}_i$, assume that $R \triangleq |\Gamma| \leq M$. We hope that different participating algorithms work on different principles and the support-set estimated by each participating algorithm includes a partially correct information about the true support-set \mathcal{T} . It may be also observed that the union of the estimated support-sets, Γ , is richer in terms of the true atoms as compared to the support-set estimated by any participating algorithm. Also note that, once the support-set is estimated, the non-zero magnitudes of \mathbf{X} can be estimated by solving a Least-Squares (LS) problem on an over-determined system of linear equations. Hence if we can identify all the true atoms included in the joint support-set Γ , we can achieve a better sparse signal estimate.

Since we are estimating the support atoms only from Γ , we need to only solve the following problem which is lower dimensional as compared to the original problem (3):

$$\mathbf{B} = \mathbf{A}_\Gamma \mathbf{X}_{\Gamma,:} + \tilde{\mathbf{W}}, \quad (4)$$

where \mathbf{A}_Γ denotes the sub-matrix formed by the columns of \mathbf{A} whose indices are listed in Γ , $\mathbf{X}_{\Gamma,:}$ denotes the submatrix formed by the rows of \mathbf{X} whose indices are listed in Γ , and $\tilde{\mathbf{W}} = \mathbf{W} + \mathbf{A}_{\Gamma^c} \mathbf{X}_{\Gamma^c,:}$. The matrix equation (4) represents a system of L linear equations which are over-determined in nature. We use the method of LS to find an approximate solution to the overdetermined system of equations in (4). Let $\mathbf{V}_{\Gamma,:}$ denote the LS solution of (4). We choose the support-set estimate of MMV-FACS as the support of \mathbf{V}^K , i.e., indices of those rows having the largest ℓ_2 -norm. Once the non-zero rows are identified, solving the resultant overdetermined solution using LS we can estimate the non-zero entries of $\hat{\mathbf{X}}$. MMV-FACS is summarized in Algorithm 1.

Algorithm 1 : MMV-FACS

Inputs: $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{B} \in \mathbb{R}^{M \times L}$, K , and $\{\hat{\mathcal{T}}_i\}_{i=1:P}$.

Assumption: $|\bigcup_{i=1}^P \hat{\mathcal{T}}_i| \leq M$.

Initialization: $\mathbf{V} = \mathbf{0} \in \mathbb{R}^{N \times 1}$.

Fusion:

- 1: $\Gamma = \bigcup_{i=1}^P \hat{\mathcal{T}}_i$;
- 2: $\mathbf{V}_{\Gamma, :} = \mathbf{A}_{\Gamma}^{\dagger} \mathbf{B}$, $\mathbf{V}_{\Gamma^c, :} = \mathbf{0}$;
- 3: $\hat{\mathcal{T}} = \text{supp}(\mathbf{V}^K)$;

Outputs: $\hat{\mathcal{T}}$ and $\hat{\mathbf{X}}$ (where $\hat{\mathbf{X}}_{\hat{\mathcal{T}}, :} = \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{B}$ and $\hat{\mathbf{X}}_{\hat{\mathcal{T}}^c, :} = \mathbf{0}$)

Remark:

An alternate approach for solving an MMV problem is to stack all the columns of \mathbf{B} to get a single measurement vector. Then (3) in a noiseless case becomes

$$\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_L \end{bmatrix}_{ML \times 1} = \begin{bmatrix} \mathbf{A} & & & \\ & \mathbf{A} & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \mathbf{A} \\ & & & & \mathbf{A} \end{bmatrix}_{ML \times NL} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_L \end{bmatrix}_{NL \times 1},$$

where \mathbf{b}_i and \mathbf{x}_i ($i = 1, 2, \dots, L$) denote the i^{th} column of \mathbf{B} and \mathbf{X} respectively. Now, we have the following SMV problem.

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{NL} \end{bmatrix}_{ML \times 1} = \begin{bmatrix} \mathbf{A} & & & \\ & \mathbf{A} & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \mathbf{A} \\ & & & & \mathbf{A} \end{bmatrix}_{ML \times NL} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{NL} \end{bmatrix}_{NL \times 1} \quad (5)$$

In principle, we can solve (5) using FACS with sparsity level LK . Note that, after stacking \mathbf{X} column-wise, we lost the joint sparsity constraint imposed on \mathbf{X} in the MMV problem in (3). The LK non-zero elements estimated from (5) using FACS can be from more than K different rows

of \mathbf{X} . In the worst case, the estimate of FACS may include non-zero elements from $\min(LK, M)$ different rows of \mathbf{X} . Then we will end up with an estimate of \mathbf{X} with LK non-zero rows, which is highly undesirable. Hence stacking the columns of the observation matrix \mathbf{B} and solving it using FACS is not advisable. Note that Step 3 in Algorithm 1 ensures that MMV-FACS estimates only K non-zero rows of \mathbf{X} .

IV. THEORETICAL STUDIES OF MMV-FACS

In this section, we will theoretically analyse the performance of MMV-FACS. We consider the general case for an arbitrary signal matrix. We also study the average case performance of MMV-FACS subsequently.

The performance analysis is characterized by SRER extended for MMV which is defined as

$$\text{SRER} \triangleq \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2}, \quad (6)$$

where \mathbf{X} and $\hat{\mathbf{X}}$ denote the actual and reconstructed signal matrix respectively.

Lemma 1. *Suppose that \mathbf{A} satisfies the relation, for some constant $\delta_{R+K} \in (0, 1)$,*

$$\|\mathbf{A}\mathbf{X}\|_F \leq \sqrt{1 + \delta_{R+K}} \|\mathbf{X}\|_F,$$

where $\|\mathbf{X}\|_0 \leq R + K$ and $\delta_{R+K} \in (0, 1)$. Here $\|\mathbf{X}\|_0$ denotes the number of non-zero rows of the matrix \mathbf{X} . Then, for every matrix \mathbf{X} ,

$$\|\mathbf{A}\mathbf{X}\|_F \leq \sqrt{1 + \delta_{R+K}} \left[\|\mathbf{X}\|_F + \frac{1}{\sqrt{R+K}} \|\mathbf{X}\|_{2,1} \right]$$

Proof: Proof is given in Appendix A.

Lemma 2. *Consider $\mathbf{A} \in \mathbb{R}^{M \times N}$ and let \mathcal{T}_1 & \mathcal{T}_2 be two subsets of $\{1, 2, \dots, N\}$ such that $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$. Assume that $\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|} \leq 1$ and let \mathbf{Y} be any matrix, such that $\text{span}(\mathbf{Y}) \in \text{span}(\mathbf{A}_{\mathcal{T}_1})$ and $\mathbf{R} = \mathbf{Y} - \mathbf{A}_{\mathcal{T}_2} \mathbf{A}_{\mathcal{T}_2}^\dagger \mathbf{Y}$. Then we have*

$$\left(1 - \frac{\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}{1 - \delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}} \right) \|\mathbf{Y}\|_2 \leq \|\mathbf{R}\|_2 \leq \|\mathbf{Y}\|_2.$$

Proof: Proof is given in Appendix A.

A. Performance Analysis for Arbitrary Signals under Measurement Perturbations

We analyse the performance of MMV-FACS for arbitrary signals and give an upper bound on the reconstruction error in Theorem 1. We also derive a sufficient condition to get an improved performance of MMV-FACS scheme over any given participating algorithm.

Theorem 1. *Let \mathbf{X} be an arbitrary signal with $\mathcal{T} = \text{supp}(\mathbf{X}^K)$. Consider the MMV-FACS setup discussed in Section III, and assume that the measurement matrix \mathbf{A} satisfies RIP with RIC δ_{R+K} . We have the following results:*

i) *Upper bound on reconstruction error: We have,*

$$\|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq C_1 \|\mathbf{X} - \mathbf{X}^K\|_F + C_2 \|\mathbf{X} - \mathbf{X}^K\|_{2,1} + C_3 \|\mathbf{X}_{\Gamma^c, :}\|_F + \nu \|\mathbf{W}\|_F$$

where $C_1 = (1 + \nu\sqrt{1 + \delta_{R+K}})$, $C_2 = \frac{\nu\sqrt{1 + \delta_{R+K}}}{\sqrt{R+K}}$, $C_3 = \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2}$, and $\nu = \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2}$.

ii) *SRER gain:*

For $\|\mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F \neq 0$ and $\|\mathbf{X}_{\Gamma^c, :}\|_F \neq 0$, MMV-FACS provides at least SRER gain of

$$\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i} \right)^2 \text{ over the } i^{\text{th}} \text{ participating algorithm if}$$

$$\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)}, \text{ where } \eta_i = \frac{\|\mathbf{X}_{\Gamma^c, :}\|_F}{\|\mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F}, \zeta = \frac{\|\mathbf{W}\|_F}{\|\mathbf{X}_{\Gamma^c, :}\|_F}, \text{ and}$$

$$\xi = \left(3\sqrt{1 + \delta_{R+K}} + 1 \right) \frac{\|\mathbf{X} - \mathbf{X}^K\|_F}{3\|\mathbf{X}_{\Gamma^c, :}\|_F} + \frac{\sqrt{1 + \delta_{R+K}}}{\sqrt{R+K}} \frac{\|\mathbf{X} - \mathbf{X}^K\|_{2,1}}{\|\mathbf{X}_{\Gamma^c, :}\|_F}.$$

Proof:

i) We have,

$$\|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq \|\mathbf{X} - \mathbf{X}^K\|_F + \|\mathbf{X}^K - \hat{\mathbf{X}}\|_F \quad (7)$$

Consider,

$$\begin{aligned} \|\mathbf{X}^K - \hat{\mathbf{X}}\|_F &\leq \|(\mathbf{X}^K)_{\hat{\mathcal{T}}, :} - \hat{\mathbf{X}}_{\hat{\mathcal{T}}, :}\|_F + \|(\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} - \hat{\mathbf{X}}_{\hat{\mathcal{T}}^c, :}\|_F \\ &\leq \|(\mathbf{X}^K)_{\hat{\mathcal{T}}, :} - \hat{\mathbf{X}}_{\hat{\mathcal{T}}, :}\|_F + \|(\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :}\|_F \quad (\because \hat{\mathbf{X}}_{\hat{\mathcal{T}}^c, :} = 0) \end{aligned} \quad (8)$$

Using the relations $\hat{\mathbf{X}}_{\hat{\mathcal{T}},:} = \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{B}$ (from Algorithm 1) and $\mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}_{\hat{\mathcal{T}}} = \mathbf{I}$, we get

$$\begin{aligned}
& \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \hat{\mathbf{X}}_{\hat{\mathcal{T}},:} \right\|_F \\
&= \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{B} \right\|_F \\
&= \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \mathbf{A}_{\hat{\mathcal{T}}}^\dagger (\mathbf{A}\mathbf{X} + \mathbf{W}) \right\|_F \quad (\because \mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}) \\
&= \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \mathbf{A}_{\hat{\mathcal{T}}}^\dagger (\mathbf{A}\mathbf{X}^K + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W}) \right\|_F \\
&= \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \left[\mathbf{A}_{\hat{\mathcal{T}}}(\mathbf{X}^K)_{\hat{\mathcal{T}},:} + \mathbf{A}_{\hat{\mathcal{T}}^c}(\mathbf{X}^K)_{\hat{\mathcal{T}}^c,:} + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W} \right] \right\|_F \\
&= \left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}_{\hat{\mathcal{T}}^c}(\mathbf{X}^K)_{\hat{\mathcal{T}}^c,:} + \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{W} \right\|_F \\
&\leq \left\| (\mathbf{A}_{\hat{\mathcal{T}}}^H \mathbf{A}_{\hat{\mathcal{T}}})^{-1} \mathbf{A}_{\hat{\mathcal{T}}}^H \mathbf{A}_{\hat{\mathcal{T}}^c}(\mathbf{X}^K)_{\hat{\mathcal{T}}^c,:} \right\|_F + \left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K) \right\|_F + \left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{W} \right\|_F \quad (9)
\end{aligned}$$

Let $\mathbf{x}^{(i)}$ denote the i^{th} column of matrix \mathbf{X} and $\mathbf{w}^{(i)}$ denote the i^{th} column of matrix \mathbf{W} , $i = 1, 2, \dots, L$. Now from Proposition 3.1 and Corollary 3.3 of [16] we obtain the following relations.

$$\left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{w}^{(i)} \right\|_2 \leq \frac{\|\mathbf{w}^{(i)}\|_2}{\sqrt{1 - \delta_{R+K}}} \quad (10)$$

$$\left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A} (\mathbf{x}^{(i)} - (\mathbf{x}^{(i)})^K) \right\|_2 \leq \frac{\|\mathbf{A} (\mathbf{x}^{(i)} - (\mathbf{x}^{(i)})^K)\|_2}{\sqrt{1 - \delta_{R+K}}} \quad (11)$$

$$\left\| (\mathbf{A}_{\hat{\mathcal{T}}}^H \mathbf{A}_{\hat{\mathcal{T}}})^{-1} \mathbf{A}_{\hat{\mathcal{T}}}^H \mathbf{A}_{\hat{\mathcal{T}}^c} ((\mathbf{x}^{(i)})^K)_{\hat{\mathcal{T}}^c} \right\|_2 \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|((\mathbf{x}^{(i)})^K)_{\hat{\mathcal{T}}^c}\|_2 \quad (12)$$

Consider (10), we get

$$\left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{w}^{(i)} \right\|_2^2 \leq \frac{\|\mathbf{w}^{(i)}\|_2^2}{1 - \delta_{R+K}} \quad \forall i = 1, 2, \dots, L$$

Summing the above equation over $i = 1, 2, \dots, L$, we obtain

$$\begin{aligned}
\sum_{i=1}^L \left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{w}^{(i)} \right\|_2^2 &\leq \frac{1}{1 - \delta_{R+K}} \sum_{i=1}^L \|\mathbf{w}^{(i)}\|_2^2 \\
\left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{W} \right\|_F^2 &\leq \frac{1}{1 - \delta_{R+K}} \|\mathbf{W}\|_F^2 \\
\left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{W} \right\|_F &\leq \frac{1}{\sqrt{1 - \delta_{R+K}}} \|\mathbf{W}\|_F. \quad (13)
\end{aligned}$$

Similarly, summing the relations in (11) and (12), we obtain

$$\left\| \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A} (\mathbf{X} - \mathbf{X}^K) \right\|_F \leq \frac{\|\mathbf{A} (\mathbf{X} - \mathbf{X}^K)\|_F}{\sqrt{1 - \delta_{R+K}}} \quad (14)$$

$$\left\| (\mathbf{A}_{\hat{\mathcal{T}}}^H \mathbf{A}_{\hat{\mathcal{T}}})^{-1} \mathbf{A}_{\hat{\mathcal{T}}}^H \mathbf{A}_{\hat{\mathcal{T}}^c} (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F \quad (15)$$

Substituting (13),(14) and (15) in (9), we get

$$\begin{aligned} \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}, :} - \hat{\mathbf{X}}_{\hat{\mathcal{T}}, :} \right\|_F &\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F + \frac{\left\| \mathbf{A} (\mathbf{X} - \mathbf{X}^K) \right\|_F}{\sqrt{1 - \delta_{R+K}}} + \frac{\left\| \mathbf{W} \right\|_F}{\sqrt{1 - \delta_{R+K}}} \\ &\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F + \frac{1}{1 - \delta_{R+K}} (\left\| \mathbf{A} (\mathbf{X} - \mathbf{X}^K) \right\|_F + \left\| \mathbf{W} \right\|_F) \end{aligned} \quad (16)$$

Substituting (16) in (8), we get

$$\left\| \mathbf{X}^K - \hat{\mathbf{X}} \right\|_F \leq \frac{1}{1 - \delta_{R+K}} \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F + \frac{1}{1 - \delta_{R+K}} (\left\| \mathbf{A} (\mathbf{X} - \mathbf{X}^K) \right\|_F + \left\| \mathbf{W} \right\|_F) \quad (17)$$

Next, we will find an upper bound for $\left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F$.

Define $\hat{\mathcal{T}}_\Delta \triangleq \Gamma \setminus \hat{\mathcal{T}}$. That is, $\hat{\mathcal{T}}_\Delta$ is the set formed by the atoms in Γ which are discarded by Algorithm 1. Since $\hat{\mathcal{T}} \subset \Gamma$, we have $\hat{\mathcal{T}}^c = \Gamma^c \cup \hat{\mathcal{T}}_\Delta$ and hence we obtain

$$\left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :} \right\|_F \leq \left\| (\mathbf{X}^K)_{\Gamma^c, :} \right\|_F + \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}_\Delta, :} \right\|_F \quad (18)$$

We also have,

$$\begin{aligned} \left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}_\Delta, :} \right\|_F &\leq \left\| (\mathbf{V}_{\Gamma, :})_{\hat{\mathcal{T}}_\Delta, :} \right\|_F + \left\| (\mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :})_{\hat{\mathcal{T}}_\Delta, :} \right\|_F \\ &\leq \left\| (\mathbf{V}_{\Gamma, :})_{\hat{\mathcal{T}}_\Delta, :} \right\|_F + \left\| \mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :} \right\|_F \end{aligned} \quad (19)$$

Note that $(\mathbf{V}_{\Gamma, :})_{\hat{\mathcal{T}}_\Delta, :}$ contains the K -rows of $\mathbf{V}_{\Gamma, :}$ with highest row ℓ_2 -norm. Therefore, using $|\hat{\mathcal{T}}| = |\mathcal{T}| = K$, we get

$$\begin{aligned} \left\| (\mathbf{V}_{\Gamma, :})_{\hat{\mathcal{T}}_\Delta, :} \right\|_F &\leq \left\| (\mathbf{V}_{\Gamma, :})_{\Gamma \setminus \mathcal{T}, :} \right\|_F \\ &= \left\| \mathbf{V}_{\Gamma \setminus \mathcal{T}, :} - (\mathbf{X}^K)_{\Gamma \setminus \mathcal{T}, :} \right\|_F \quad (\because (\mathbf{X}^K)_{\Gamma \setminus \mathcal{T}, :} = 0) \\ &\leq \left\| (\mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :}) \right\|_F. \end{aligned} \quad (20)$$

Substituting (20) in (19), we get

$$\left\| (\mathbf{X}^K)_{\hat{\mathcal{T}}_\Delta, :} \right\|_F \leq 2 \left\| (\mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :}) \right\|_F. \quad (21)$$

Now, consider

$$\begin{aligned}
& \|\mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger \mathbf{B} - (\mathbf{X}^K)_{\Gamma, :}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger (\mathbf{A}\mathbf{X} + \mathbf{W}) - (\mathbf{X}^K)_{\Gamma, :}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger (\mathbf{A}\mathbf{X}^K + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W}) - (\mathbf{X}^K)_{\Gamma, :}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger (\mathbf{A}_\Gamma (\mathbf{X}^K)_{\Gamma, :} + \mathbf{A}_{\Gamma^c} (\mathbf{X}^K)_{\Gamma^c, :} + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W}) - (\mathbf{X}^K)_{\Gamma, :}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger \mathbf{A}_{\Gamma^c} (\mathbf{X}^K)_{\Gamma^c, :} + \mathbf{A}_\Gamma^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{A}_\Gamma^\dagger \mathbf{W}\|_F \quad (\because \mathbf{A}_\Gamma^\dagger \mathbf{A}_\Gamma = \mathbf{I}) \\
&\leq \|\mathbf{A}_\Gamma^\dagger \mathbf{A}_{\Gamma^c} (\mathbf{X}^K)_{\Gamma^c, :}\|_F + \|\mathbf{A}_\Gamma^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{A}_\Gamma^\dagger \mathbf{W}\|_F.
\end{aligned} \tag{22}$$

Using (13), (14) and (15) in (22), we get

$$\begin{aligned}
\|\mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :}\|_F &\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\Gamma^c, :}\|_F + \frac{\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F}{\sqrt{1 - \delta_{R+K}}} + \frac{\|\mathbf{W}\|_F}{\sqrt{1 - \delta_{R+K}}} \\
&\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\Gamma^c, :}\|_F + \frac{\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F}{1 - \delta_{R+K}} + \frac{\|\mathbf{W}\|_F}{1 - \delta_{R+K}}. \\
&\quad (\because 0 < 1 - \delta_{R+K} < 1)
\end{aligned} \tag{23}$$

Using (21) and (23) in (18), we get

$$\|(\mathbf{X}^K)_{\hat{\mathcal{T}}^c, :}\|_F \leq \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\Gamma^c, :}\|_F + \frac{2}{1 - \delta_{R+K}} (\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{W}\|_F). \tag{24}$$

Let $\mathbf{x}_1^{(i)}$ denote the i^{th} column of matrix \mathbf{X}^K . The, we have,

$$\begin{aligned}
& \|\mathbf{A}\mathbf{x}_1^{(i)}\|_2^2 \leq (1 + \delta_{R+K}) \|\mathbf{x}_1^{(i)}\|_2^2 \quad (\because \mathbf{A} \text{ satisfies RIP}) \\
& \sum_{i=1}^L \|\mathbf{A}\mathbf{x}_1^{(i)}\|_2^2 \leq \sum_{i=1}^L (1 + \delta_{R+K}) \|\mathbf{x}_1^{(i)}\|_2^2 \\
& \|\mathbf{A}\mathbf{X}^K\|_F^2 \leq (1 + \delta_{R+K}) \|\mathbf{X}^K\|_F^2
\end{aligned} \tag{25}$$

Using Lemma 1 and (25), we get

$$\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F \leq \sqrt{1 + \delta_{R+K}} \left[\|\mathbf{X} - \mathbf{X}^K\|_F + \frac{1}{\sqrt{R+K}} \|(\mathbf{X} - \mathbf{X}^K)\|_{2,1} \right] \tag{26}$$

Substituting (24) in (17), we get

$$\begin{aligned}
\|\mathbf{X}^K - \hat{\mathbf{X}}\|_F &\leq \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|(\mathbf{X}^K)_{\Gamma^c, :}\|_F + \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2} [\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{W}\|_F] \\
&\leq \nu \sqrt{1 + \delta_{R+K}} \|\mathbf{X} - \mathbf{X}^K\|_F + \frac{\nu \sqrt{1 + \delta_{R+K}} \|\mathbf{X} - \mathbf{X}^K\|_{2,1}}{\sqrt{R + K}} \\
&\quad + \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|\mathbf{X}_{\Gamma^c, :}\|_F + \nu \|\mathbf{W}\|_F, \quad (\text{using (26)}) \tag{27}
\end{aligned}$$

where $\nu = \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2}$.

Substituting (27) in (7) and using the definitions of C_1, C_2 , and C_3 , we get

$$\|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq C_1 \|\mathbf{X} - \mathbf{X}^K\|_F + C_2 \|\mathbf{X} - \mathbf{X}^K\|_{2,1} + C_3 \|\mathbf{X}_{\Gamma^c, :}\|_F + \nu \|\mathbf{W}\|_F. \tag{28}$$

ii) Using (28) and the definitions of ξ and η_i , we get

$$\begin{aligned}
\|\mathbf{X} - \hat{\mathbf{X}}\|_F &\leq \frac{1 + \delta_{R+K} + 3\zeta + 3\xi}{(1 - \delta_{R+K})^2} \|\mathbf{X}_{\Gamma^c, :}\|_F \\
&= \frac{1 + \delta_{R+K} + 3\zeta + 3\xi}{(1 - \delta_{R+K})^2} \eta_i \left\| (\mathbf{X} - \hat{\mathbf{X}}_i)_{\hat{\mathcal{T}}_i^c, :} \right\|_F \quad (\because (\hat{\mathbf{X}}_i)_{\hat{\mathcal{T}}_i^c, :} = \mathbf{0}) \\
&\leq \frac{1 + \delta_{R+K} + 3\zeta + 3\xi}{(1 - \delta_{R+K})^2} \eta_i \|\mathbf{X} - \hat{\mathbf{X}}_i\|_F.
\end{aligned}$$

Hence, we obtain the relation for SRER for MMV-FACS, in case of arbitrary signals, as

$$\begin{aligned}
\text{SRER}|_{\text{MMV-FACS}} &= \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2} \\
&\geq \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}_i\|_F^2} \times \left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i} \right)^2
\end{aligned}$$

Hence MMV-FACS provides at least SRER gain of $\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i} \right)^2$ over i^{th} algorithm if $\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)}$.

Note that $\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)} < 1$. □

B. Exactly K -sparse Matrix

Theorem 1 considered the case when \mathbf{X} is an arbitrary matrix. If \mathbf{X} is a K -sparse matrix then we have $\mathbf{X} = \mathbf{X}^K$ and $\xi = 0$. Thus, it follows from Theorem 1 that, MMV-FACS

provides at least SRER gain of $\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)\eta_i}\right)^2$ over i^{th} participating algorithm if $\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)}$. Thus, the improvement in the SRER gain provided by MMV-FACS over the i^{th} Algorithm for a K -sparse matrix is greater than that of an arbitrary matrix by a factor of $\left(1 + \frac{3\xi}{(1 + \delta_{R+K} + 3\zeta)}\right)^2$.

The second part of Theorem 1 considers the case when $\|\mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F \neq 0$ and $\|\mathbf{X}_{\Gamma^c, :}\|_F \neq 0$. If $\|\mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F \neq 0$, then $\hat{\mathcal{T}}_i \not\subseteq \mathcal{T}$. Also, $\|\mathbf{X}_{\Gamma^c, :}\|_F = 0$ implies $\mathcal{T} \subseteq \Gamma$. Suppose $\|\mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F = 0$, then the support-set is correctly estimated by i^{th} algorithm and further performance improvement is not possible by MMV-FACS. Hence we consider the case where $\|\mathbf{X}_{\Gamma^c, :}\|_F = 0$, and derive the condition for exact reconstruction by MMV-FACS in the following proposition.

Proposition 1. *Assume that $\|\mathbf{X}_{\Gamma^c, :}\|_F = 0$ and all other conditions in Theorem 1 hold good. Then, in clean measurement case ($\mathbf{W} = \mathbf{0}$), MMV-FACS estimates the support-set correctly and provides exact reconstruction.*

Proof: We have

$$\mathbf{X}_{\Gamma^c, :} = \mathbf{0} \Rightarrow \mathcal{T} \subset \Gamma \quad (29)$$

$$\mathbf{B} = \mathbf{A}_{\mathcal{T}}\mathbf{X}_{\hat{\mathcal{T}}, :} + \mathbf{W} \quad (30)$$

From Algorithm 1, we have $\mathbf{V} \in \mathbb{R}^{N \times L}$ where $\mathbf{V}_{\Gamma^c, :} = \mathbf{0}$, and

$$\begin{aligned} \mathbf{V}_{\Gamma, :} &= \mathbf{A}_{\Gamma}^{\dagger} \mathbf{B} \\ &= \mathbf{A}_{\Gamma}^{\dagger} (\mathbf{A}_{\mathcal{T}} \mathbf{X}_{\hat{\mathcal{T}}, :} + \mathbf{W}) && \text{(using (30))} \\ &= \mathbf{A}_{\Gamma}^{\dagger} (\mathbf{A}_{\Gamma} \mathbf{X}_{\Gamma, :} + \mathbf{W}) && \text{(using (29))} \\ &= \mathbf{X}_{\Gamma, :} + \mathbf{A}_{\Gamma}^{\dagger} \mathbf{W}. \end{aligned}$$

If $\mathbf{W} = \mathbf{0}$, then $\mathbf{V}_{\Gamma, :} = \mathbf{X}_{\Gamma, :}$ and $\mathbf{V} = \mathbf{X}$ ($\because \mathcal{T} \subset \Gamma$). Thus MMV-FACS estimates the support-set correctly from \mathbf{V} . \square

In practice, the original signal is not known and hence it is not possible to evaluate the performance w.r.t. the true signal. Hence in applications, the decrease in energy of the residual is often treated as a measure of performance improvement. Proposition 2 gives a sufficient

condition for decrease in the energy of the residual matrix obtained by MMV-FACS over the i^{th} participating algorithm.

Proposition 2. For a K -sparse matrix \mathbf{X} , let $\mathbf{R} = \mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{B}$ and $\mathbf{R}_i = \mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{B}$ represent the residue matrix of MMV-FACS and i^{th} Algorithm respectively. Assume that $\frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}}(1 + \delta_{R+K} + 3\zeta) \leq \left(\frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta \right)$ is satisfied then we have, $\|\mathbf{R}\|_F \leq \|\mathbf{R}_i\|_F$.

Proof: We have,

$$\begin{aligned}
\|\mathbf{R}\|_F &= \left\| \mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{B} \right\|_F \\
&= \left\| \mathbf{A}\mathbf{X} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger (\mathbf{A}\mathbf{X} + \mathbf{W}) \right\|_F \quad (\because \mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}) \\
&= \left\| \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{X}_{\hat{\mathcal{T}},:} + \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger (\mathbf{A}_{\hat{\mathcal{T}}} \mathbf{X}_{\hat{\mathcal{T}},:} + \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} + \mathbf{W}) \right\|_F \\
&= \left\| \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{W} \right\|_F \quad (\because \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}_{\hat{\mathcal{T}}} = \mathbf{I}) \\
&\leq \left\| \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} \right\|_F + \left\| \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^\dagger \mathbf{W} \right\|_F \\
&\leq \left\| \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c,:} \right\|_F + \|\mathbf{W}\|_F \quad (\text{Using Lemma 2}) \\
&= \left\| \mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}},:} \right\|_F + \|\mathbf{W}\|_F \quad (\because \mathcal{T} = \text{supp}(\mathbf{X})) \\
&\leq \sqrt{1 + \delta_{R+K}} \left\| \mathbf{X}_{\hat{\mathcal{T}}^c,:} \right\|_F + \|\mathbf{W}\|_F \quad \left(\because |\mathcal{T} \setminus \hat{\mathcal{T}}| \leq K \quad \& \quad \delta_K \leq \delta_{R+K} \right).
\end{aligned}$$

Using (24) we have,

$$\begin{aligned}
\|\mathbf{R}\|_F &\leq \sqrt{1 + \delta_{R+K}} \left(\frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \|\mathbf{X}_{\Gamma^c,:}\|_F + \frac{2}{1 - \delta_{R+K}} \|\mathbf{W}\|_F \right) + \|\mathbf{W}\|_F \\
&\leq \frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} \left(1 + \delta_{R+K} + 2\zeta + \frac{1 - \delta_{R+K}}{\sqrt{1 + \delta_{R+K}}} \zeta \right) \|\mathbf{X}_{\Gamma^c,:}\|_F \\
&\leq \frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \|\mathbf{X}_{\Gamma^c,:}\|_F.
\end{aligned} \tag{31}$$

Now, consider

$$\begin{aligned}
\|\mathbf{R}_i\|_F &= \left\| \mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{B} \right\|_F \\
&= \left\| \mathbf{A}\mathbf{X} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger (\mathbf{A}\mathbf{X} + \mathbf{W}) \right\|_F \quad (\because \mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}) \\
&= \left\| \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \left(\mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} + \mathbf{W} \right) \right\|_F \\
&= \left\| \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{W} \right\|_F \\
&\geq \left\| \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} \right\|_F - \left\| \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{W} \right\|_F \\
&\quad (\text{Using reverse triangle inequality}) \\
&= \left\| \mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i, :} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i, :} \right\|_F - \left\| \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^\dagger \mathbf{W} \right\|_F \\
&\geq \left(1 - \frac{\delta_{R+K}}{1 - \delta_{R+K}} \right) \left\| \mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i, :} \right\|_F - \|\mathbf{W}\|_F \\
&\quad (\text{Using Lemma 2 \& } \delta_{2K} \leq \delta_{R+K}) \\
&\geq \frac{1 - 2\delta_{R+K}}{1 - \delta_{R+K}} \sqrt{1 - \delta_{R+K}} \left\| \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} \right\|_F - \|\mathbf{W}\|_F \quad (\because |\mathcal{T} \setminus \hat{\mathcal{T}}| \leq K \quad \& \quad \delta_K \leq \delta_{R+K}) \\
&= \left(\frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta \right) \left\| \mathbf{X}_{\Gamma^c, :} \right\|_F. \tag{32}
\end{aligned}$$

From (31) and (32) we get a sufficient condition for $\|\mathbf{R}\|_F \leq \|\mathbf{R}_i\|_F$ as

$$\frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \leq \left(\frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta \right). \tag{33}$$

Thus, if (33) is satisfied, MMV-FACS produces a smaller residual matrix (in the Frobenius norm sense) than that of the i^{th} participating algorithm. ■

C. Average Case Analysis

Intuitively, we expect multiple measurement vector problem to perform better than the single measurement vector case. However, if each measurement vector is the same, i.e., in the worst case, we have $\mathbf{x}^{(i)} = \mathbf{c}$, $\forall i = 1, \dots, L$, then we do not have any additional information on \mathbf{X} than that provided by a single measurement vector $\mathbf{x}^{(1)}$. So far we have carried out only the worst case analysis, i.e., conditions under which the algorithm is able to recover any joint sparse matrix \mathbf{X} . This approach does not provide insight into the superiority of sparse signal

reconstruction with multiple measurement vectors compared to the single measurement vector case.

To notice a performance gain with multiple measurement vectors, next we proceed with an average case analysis. Here we impose a probability model on the K sparse \mathbf{X} as suggested by Remi *et al.* [17]. In particular, on the support-set \mathcal{T} , we impose that $\mathbf{X}_{\mathcal{T},:} = \Sigma\Phi$, where Σ is a $K \times K$ diagonal matrix with positive diagonal entries and Φ is a $K \times L$ random matrix with i.i.d. Gaussian entries. Our goal is to show that, under this signal model, the typical behaviour of MMV-FACS is better than in the worst case.

Theorem 2. *Consider the MMV-FACS setup discussed in Section III. Assume a Gaussian signal model, i.e., $\mathbf{X}_{\mathcal{T},:} = \Sigma\Phi$, where Σ is a $K \times K$ diagonal matrix with positive diagonal entries and Φ is a $K \times L$ random matrix with i.i.d. Gaussian entries. Let \mathbf{e}_i denote a $|\Gamma| \times 1$ vector with a '1' in the i^{th} coordinate and '0' elsewhere. Let $\eta = \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2$ and*

$$\gamma = \frac{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \frac{\eta}{C_2(L)}}{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}.$$

where $C_2(L) = \mathbb{E} \|Z\|_2$ with $Z = (Z_1, \dots, Z_L)$ being a vector of independent standard normal variables. Assume that $\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 > \frac{\eta}{C_2(L)}$. Let Θ denote the event that MMV-FACS picks all correct indices from the union-set Γ . Then, we have,

$$P(\Theta) \geq 1 - K \exp(-2A_2(L)\gamma^2),$$

where $A_2(L) = \left(\frac{\ddot{\Gamma}(\frac{L+1}{2})}{\ddot{\Gamma}(\frac{L}{2})} \right)^2 \approx \frac{L}{2}$, $\ddot{\Gamma}(\cdot)$ denotes the Gamma function.

Proof: We have,

$$\begin{aligned} P(\Theta) &= P \left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{B} \right\|_2 > \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{B} \right\|_2 \right) \\ &= P \left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger (\mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} + \mathbf{W}) \right\|_2 > \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger (\mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} + \mathbf{W}) \right\|_2 \right) \\ &> P \left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left(\left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 \right) \right. \\ &\quad \left. \geq \max_{j \in (\Gamma \setminus \mathcal{T})} \left(\left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 + \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 \right) \right) \end{aligned}$$

(Using reverse triangle inequality and triangle inequality respectively)

$$\begin{aligned}
&= P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \right. \\
&\quad \left. > \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 \right) \\
&= P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 > \eta\right) \\
&= 1 - P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq \eta\right) \\
&\geq 1 - P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq C\right) \\
&\quad - P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \geq C - \eta\right). \tag{34}
\end{aligned}$$

Now, let us derive an upper bound for $P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq C\right)$. Influenced by the concentration of measure results in [17], we set

$$C = (1 - \epsilon_1)C_2(L) \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2, \tag{35}$$

where $0 < \epsilon_1 < 1$.

Using (5.5) in [17], we get,

$$P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq C\right) \leq |\mathcal{T}| \exp(-A_2(L)\epsilon_1^2). \tag{36}$$

To bound the second probability, consider

$$\begin{aligned}
&P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \Phi \right\|_2 \geq C - \eta\right) \\
&= P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \Phi \right\|_2 \geq (C - \eta) \frac{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}\right).
\end{aligned}$$

Let

$$1 + \epsilon_2 = \frac{C - \eta}{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2} \tag{37}$$

Using equation (5.3) in [17]

$$\begin{aligned}
&P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \Phi \right\|_2 \geq (1 + \epsilon_2)C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2\right) \\
&\leq |\mathcal{T}| \exp(-A_2(L)\epsilon_2^2). \tag{38}
\end{aligned}$$

For the above inequality to hold, it is required that $\epsilon_2 > 0$. By setting $\epsilon_2 = \epsilon_1$, and using (35) and (37), we get

$$\epsilon_1 = \frac{(1 - \epsilon_1)C_2(L) \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \eta}{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2} - 1.$$

Now, solving for ϵ , we get

$$\epsilon_1 = \frac{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \frac{\eta}{C_2(L)}}{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}.$$

Clearly $\epsilon_1 < 1$ and by the assumption in the theorem $\epsilon_1 > 0$. Hence we have $0 < \epsilon_1 < 1$. Also, note that $\gamma = \epsilon_1$. Substituting (36) and (38) in (34), we get

$$P(\Theta) \geq 1 - K \exp(-2A_2(L)\gamma^2).$$

■

Since $A_2(L) \approx \frac{L}{2}$, the probability that MMV-FACS selects all correct indices from the union set increases as L increases. Thus, more than one measurement vector improves the performance.

V. NUMERICAL EXPERIMENTS AND RESULTS

We conducted numerical experiments using synthetic data and real signals to evaluate the performance of MMV-FACS. The performance is evaluated using ASRER which is defined as

$$\text{ASRER} = \frac{\sum_{j=1}^{n_{\text{trials}}} \left\| \mathbf{X}_j \right\|_F^2}{\sum_{j=1}^{n_{\text{trials}}} \left\| \mathbf{X}_j - \hat{\mathbf{X}}_j \right\|_F^2}, \quad (39)$$

where \mathbf{X}_j and $\hat{\mathbf{X}}_j$ denote the actual and reconstructed jointly sparse signal matrix in the j^{th} trial respectively, and n_{trials} denotes the total number of trials.

A. Synthetic Sparse Signals

For noisy measurement simulations, we define the SMNR as

$$\text{SMNR} \triangleq \frac{\mathbb{E}\{\left\| \mathbf{x}^{(i)} \right\|_2^2\}}{\mathbb{E}\{\left\| \mathbf{w}^{(i)} \right\|_2^2\}},$$

where $\mathbb{E}\{\cdot\}$ denotes the mathematical expectation operator. The simulation set-up is described below.

1) *Experimental Setup*: Following steps are involved in the simulation:

- i) Generate elements of $\mathbf{A}_{M \times N}$ independently from $\mathcal{N}(0, \frac{1}{M})$ and normalize each column norm to unity.
- ii) Choose K non-zero locations uniformly at random from the set $\{1, 2, \dots, N\}$ and fill those K rows of \mathbf{X} based on the choice of signal characteristics:
 - a) *Gaussian sparse signal matrix*: Non-zero values independently from $\mathcal{N}(0, 1)$.
 - b) *Rademacher sparse signal matrix*: Non-zero values are set to +1 or -1 with probability $\frac{1}{2}$.

Remaining $N - K$ rows of \mathbf{X} are set to zero.

- iii) The MMV measurement matrix \mathbf{B} is computed as $\mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}$, where the columns of \mathbf{W} , $\mathbf{w}^{(i)}$'s are independent and their elements are i.i.d. as Gaussian with variance determined from the specified SMNR.
- iv) Apply the MMV sparse recovery method.
- v) Repeat steps i-iv, S times.
- vi) Find ASRER using (39).

2) *Results and Discussions*: We used M-OMP, M-SP, M-BPDN [18], and M-FOCUSS [1] as the participating algorithms in MMV-FACS. The software code for M-BPDN was taken from SPGL1 software package [19]. Since M-FOCUSS and M-BPDN algorithms may not yield an exact K -sparse solution, we estimate the support-set as the indices of the K rows with largest ℓ_2 norm. We fixed the sparse signal dimension $N = 500$ and sparsity level $K = 20$ in the simulation the result were calculated by averaging over 1,000 trials ($S = 1,000$). We use an *oracle estimator* for performance benchmarking. The *oracle estimator* is aware of the true support-set and finds the non-zero entries of the sparse matrix by solving LS.

The empirical performance of MMV reconstruction algorithms for different values of M is shown in Fig. 1. The simulation parameters are $L = 20$, SMNR= 20 dB and \mathbf{X} is chosen as Gaussian sparse signal matrix. For $M = 35$, MMV-FACS (M-BPDN,M-FOCUSS) gave 10.67 dB and 4.27 dB improvement over M-BPDN and M-FOCUSS respectively.

Fig. 2 depicts the performance of Rademacher sparse signal matrix for different values of M where we set $L = 20$ and SMNR= 20 dB. We again observe similar performance improvement

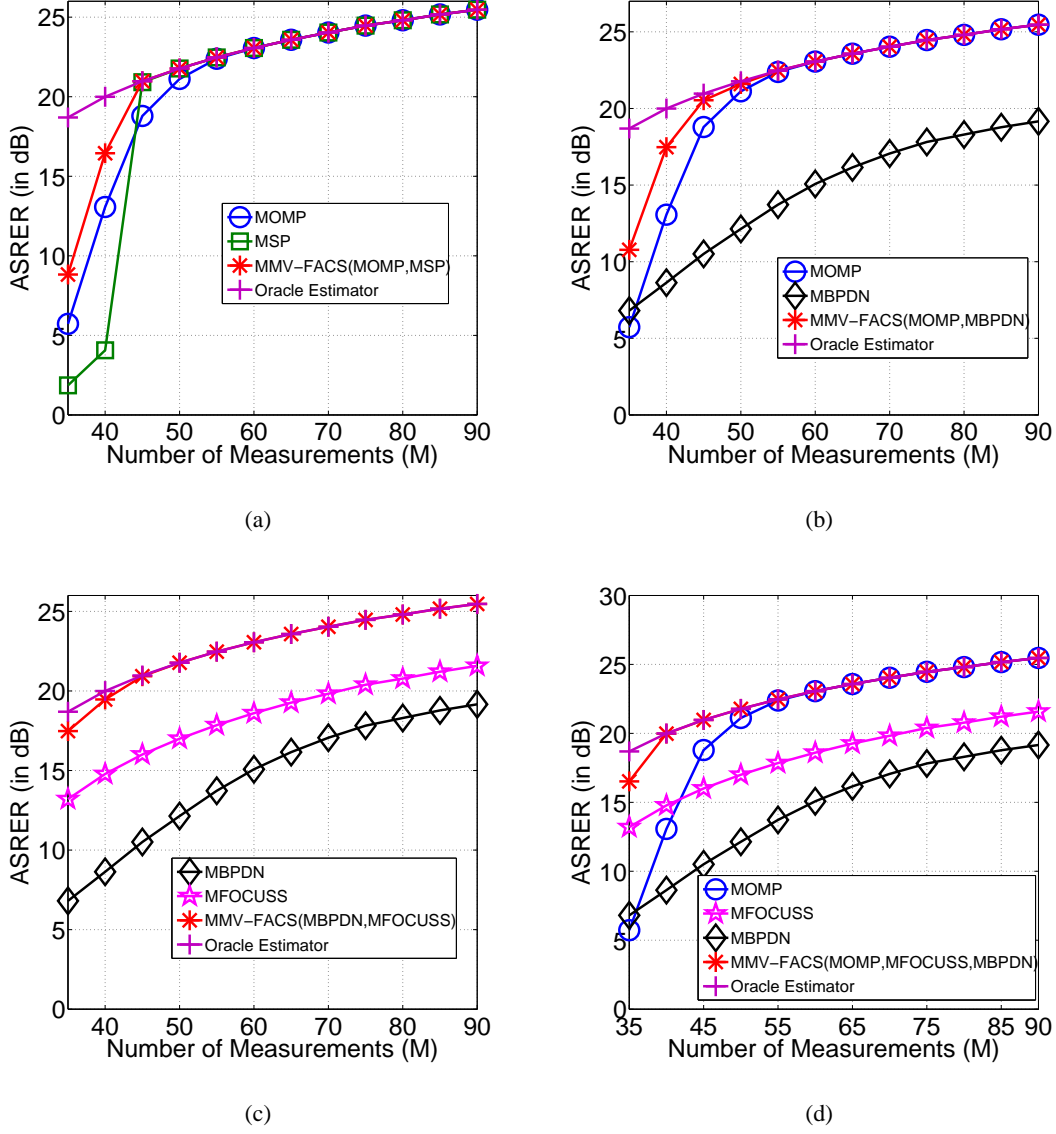


Fig. 1. Performance of MMV-FACS, averaged over 1,000 trials, for Gaussian sparse signal matrices with $\text{SMNR} = 20$ dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$, and number of measurement vectors $L = 20$.

as in the case of Gaussian sparse signal matrix. For example, for $M = 35$, MMV-FACS(MOMP,M-BPDN) showed 7.56 dB and 4.32 dB over M-OMP and M-BPDN respectively.

A comparison of MMV reconstruction techniques is shown in Fig. 3 for Gaussian sparse signal matrix for different values of L where we set $M = 50$ and $\text{SMNR} = 20$ dB. It may be observed that MMV-FACS gave a significant performance improvement over the participating algorithms. Specifically, MMV-FACS(M-OMP,M-SP) improved the performance by 5.77 dB and 4.94 dB

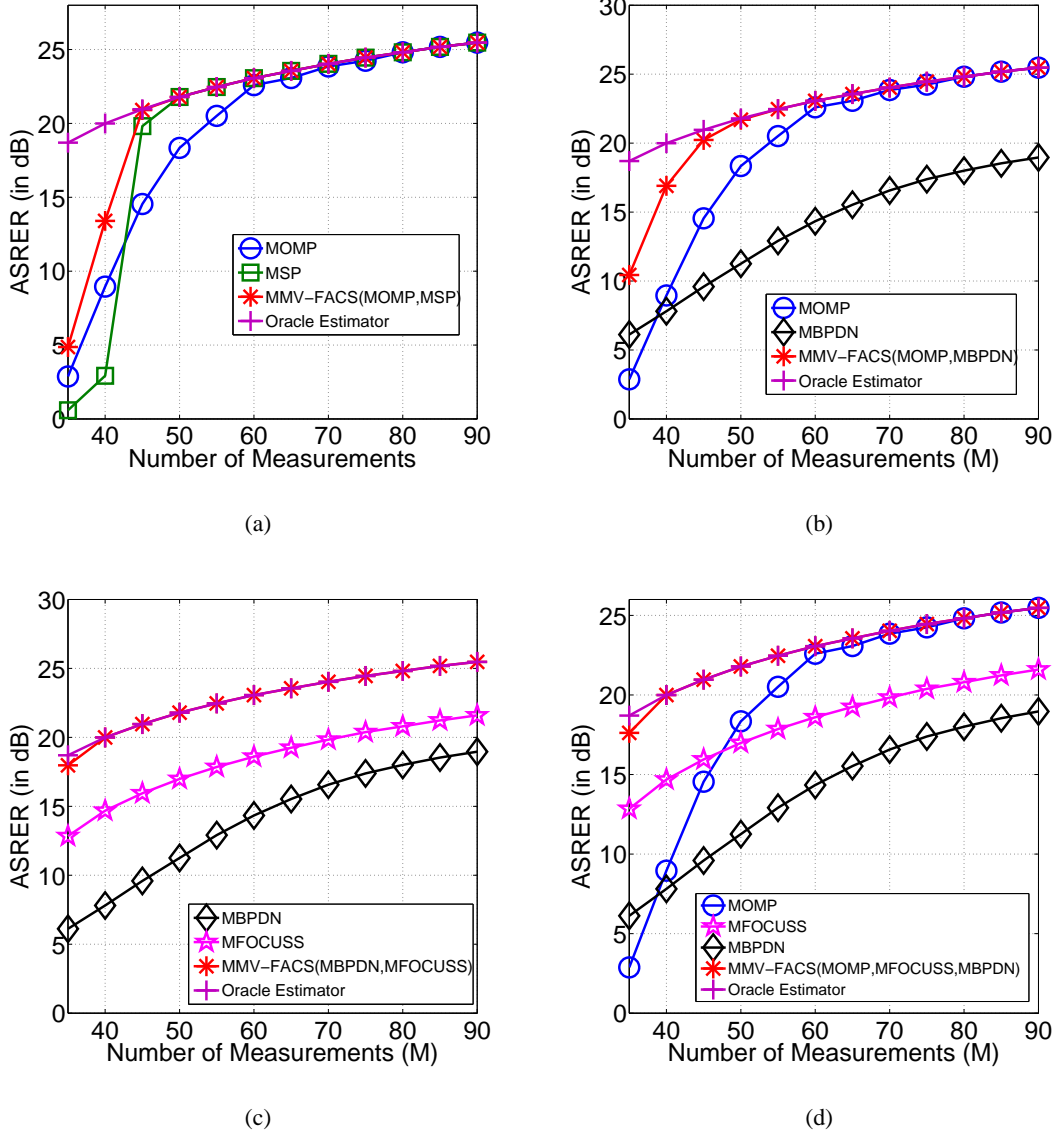


Fig. 2. Performance of MMV-FACS, averaged over 1000 trials, for Rademacher sparse signal matrices with $\text{SMNR} = 20$ dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$ and number of measurement vectors $L = 20$.

over M-OMP and M-SP respectively.

To show the dependency of recovery performance on SMNR , we conducted simulations for different values of SMNR . Fig. 4 illustrates the performance for Gaussian sparse signal matrix where $L = 10$ and $M = 45$. An additional ASRR improvement of 2.51 dB and 2.08 dB were achieved as compared to M-OMP and M-FOCUSS respectively for $\text{SMNR} = 10$ dB. This shows the robustness of MMV-FACS to noisy measurements.

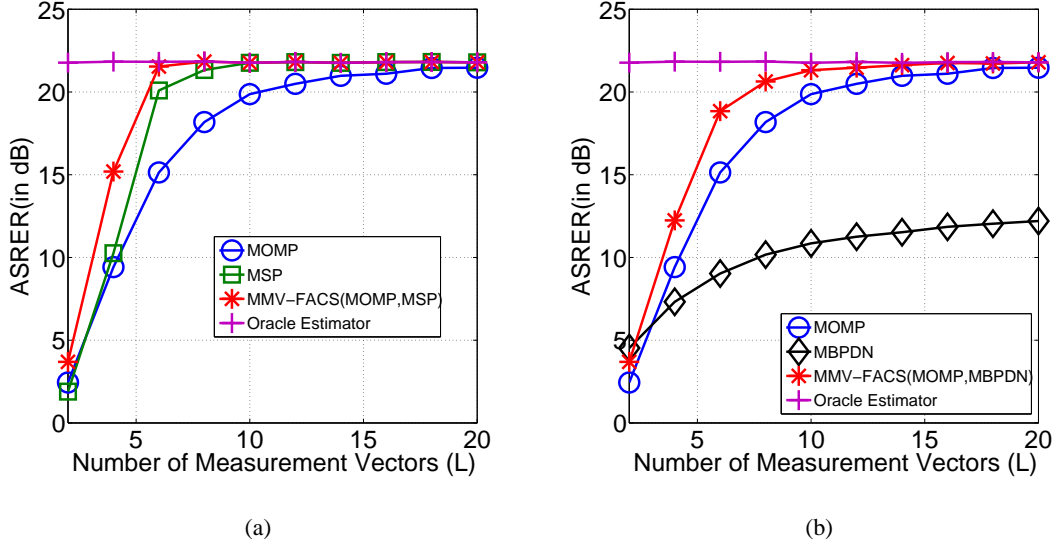


Fig. 3. Performance of MMV-FACS, averaged over 1000 trials, for Gaussian sparse signal matrices with $\text{SMNR} = 20$ dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$, and number of measurements $M = 50$.

From the above simulation results it can be seen that MMV-FACS improved the sparse signal recovery compared to participating algorithms.

3) *Reproducible Research*: We provide necessary Matlab codes to reproduce all the figures, publicly downloadable from <http://www.ece.iisc.ernet.in/~ssplab/Public/MMVFACS.tar.gz>.

B. Real Compressible Signals

To evaluate the performance of MMV-FACS on compressible signals and real world data, we used the data set ‘05091.dat’ from MIT-BIH Atrial Fibrillation Database [20]. The recording is of 10 hours in duration, and contains two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ± 10 millivolts. We selected the first 250 time points of the recording as the data set used in our experiment. We used a randomly generated Gaussian sensing matrix of size $M \times 250$, with different values of M in the experiment. We assumed sparsity level $K = 50$ and used M-OMP and M-SP as the participating algorithms. The reconstruction results are shown in Fig. 5.

Similar to synthetic signals, MMV-FACS shows a better ASRER compared to the participating algorithms M-OMP and M-SP. This demonstrates the advantage of MMV-FACS in real-life applications, requiring fewer measurement samples to yield an approximate reconstruction.

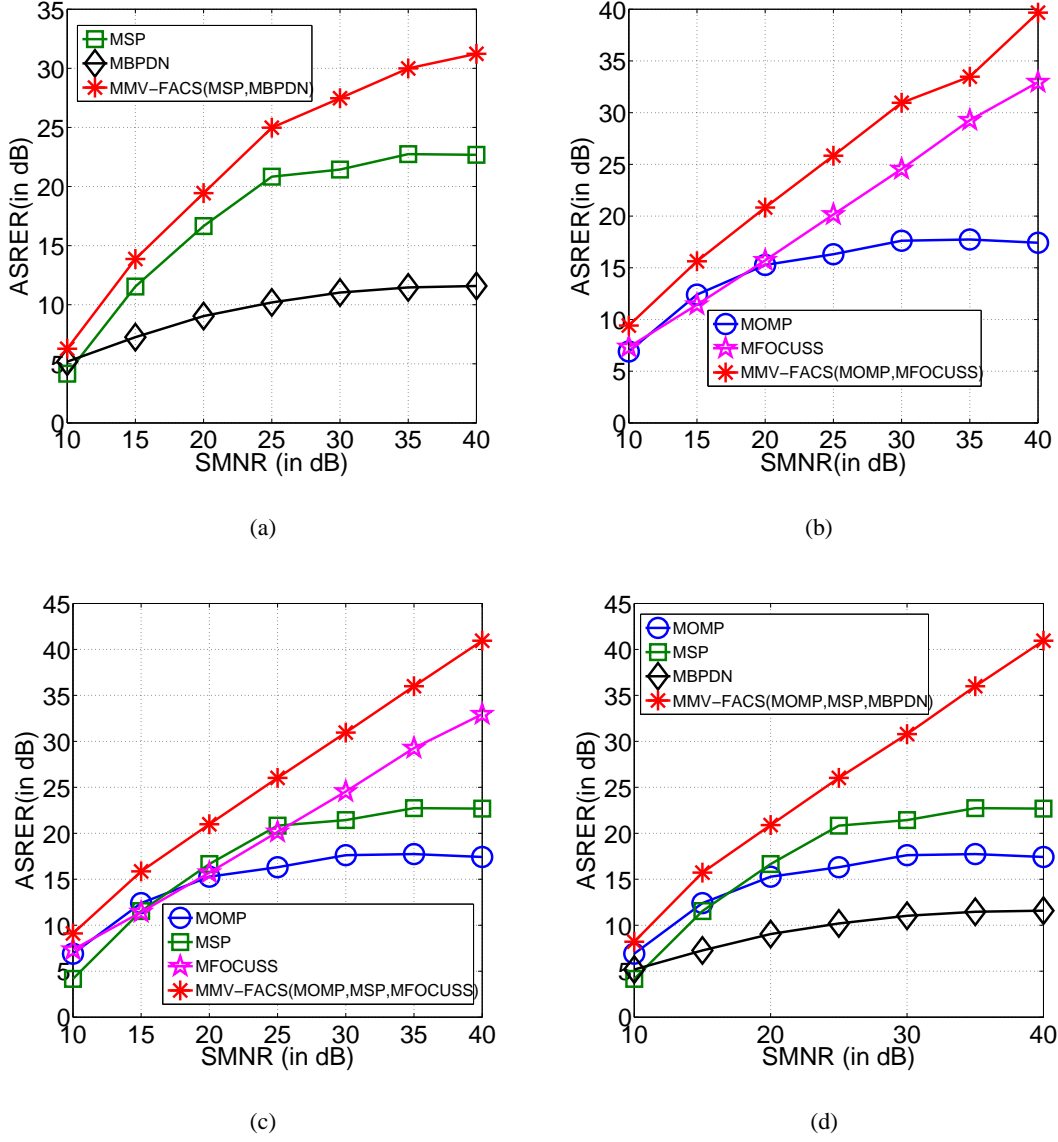


Fig. 4. Performance of MMV-FACS, averaged over 1000 trials, for Gaussian sparse signal matrices with SMNR = 20 dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$, and number of measurements $M = 45$, and number of measurement vectors $L = 10$.

VI. CONCLUSIONS

In this paper, we extended FACS to the MMV case and showed that MMV-FACS improves sparse signal matrix reconstruction. Using RIP, we theoretically analysed the proposed scheme and derived sufficient conditions for the performance improvement over the participating algorithm. Using Monte-Carlo simulations, we showed the performance improvement of the proposed

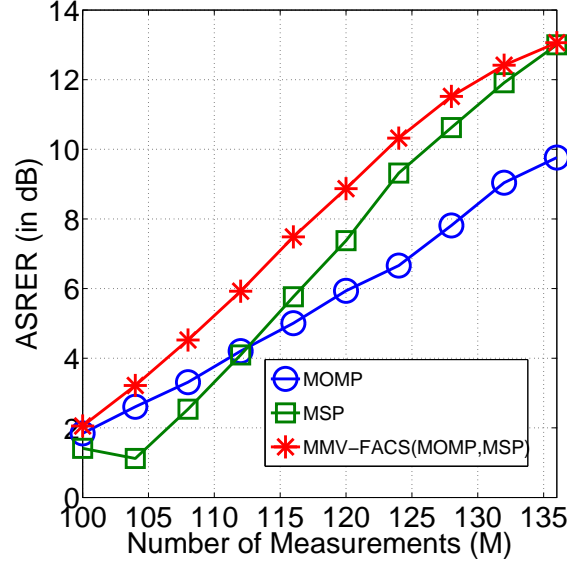


Fig. 5. Real Compressible signals: Performance of MMV-FACS for 2-channel ECG signals from MIT-BIH Atrial Fibrillation Database [20].

scheme over the participating methods. Though this paper discusses only the extension of FACS for MMV problem, a similar approach can be used to extend the other fusion algorithms developed by Ambat *et al.* [15], [21], [22].

APPENDIX

PROOF OF LEMMA 1

The proof is inspired by Proposition 3.5 by Needell and Tropp [16].

Define set S as the convex combination of all matrices which are $R + K$ sparse and have unit Frobenius norm.

$$S = \text{conv} \left\{ \mathbf{X} : \|\mathbf{X}\|_0 \leq R + K, \|\mathbf{X}\|_F = 1 \right\}$$

Using the relation $\|\mathbf{AX}\|_F \leq \sqrt{1 + \delta_{R+K}} \|\mathbf{X}\|_F$, we get,

$$\arg \max_{\mathbf{X} \in S} \|\mathbf{AX}\|_F \leq \sqrt{1 + \delta_{R+K}}.$$

Define

$$Q = \left\{ \mathbf{X} : \|\mathbf{X}\|_F + \frac{1}{\sqrt{R+K}} \|\mathbf{X}\|_{2,1} \leq 1 \right\}.$$

The lemma essentially claims that

$$\arg \max_{\mathbf{X} \in Q} \|\mathbf{A}\mathbf{X}\|_F \leq \arg \max_{\mathbf{X} \in S} \|\mathbf{A}\mathbf{X}\|_F.$$

To prove this, it is sufficient to ensure that $Q \subset S$.

Consider a matrix $\mathbf{X} \in Q$. Partition the support of \mathbf{X} into sets of size $R + K$. Let set \mathcal{I}_0 contain the indices of the $R + K$ rows of \mathbf{X} which have largest row ℓ_2 -norm, breaking ties lexicographically. Let set \mathcal{I}_1 contain the indices of the next largest (row ℓ_2 -norm) $R + K$ rows and so on. The final block \mathcal{I}_J may have lesser than $R + K$ components. This partition gives rise to the following decomposition:

$$\mathbf{X} = \mathbf{X}|_{\mathcal{I}_0} + \sum_{j=1}^J \mathbf{X}|_{\mathcal{I}_j} = \lambda_0 \mathbf{Y}_0 + \sum_{j=1}^J \lambda_j \mathbf{Y}_j,$$

where $\lambda_j = \|\mathbf{X}|_{\mathcal{I}_j}\|_F$ and $\mathbf{Y}_j = \lambda_j^{-1} \mathbf{X}|_{\mathcal{I}_j}$.

By construction each matrix \mathbf{Y}_j belongs to S because it is $R + K$ sparse and has unit Frobenius norm. We will show that $\sum_j \lambda_j \leq 1$. This implies that \mathbf{X} can be written as a convex combination of matrices from the set S . As a result $\mathbf{X} \in S$. Therefore, $Q \subset S$.

Fix some j in the range $\{1, 2, \dots, J\}$. Then, \mathcal{I}_j contains at most $R + K$ elements and \mathcal{I}_{j-1} contains exactly $R + K$ elements. Therefore,

$$\lambda_j = \|\mathbf{X}|_{\mathcal{I}_j}\|_F \leq \sqrt{R + K} \|\mathbf{X}|_{\mathcal{I}_j}\|_{2,\infty} \leq \sqrt{R + K} \cdot \frac{1}{R + K} \|\mathbf{X}|_{\mathcal{I}_{j-1}}\|_{2,1}.$$

The last inequality holds because the row ℓ_2 -norm of \mathbf{X} on the set \mathcal{I}_{j-1} dominates its largest row ℓ_2 -norm in \mathcal{I}_j . Summing these relations, we get

$$\sum_{j=1}^J \lambda_j \leq \frac{1}{\sqrt{R + K}} \sum_{j=1}^J \|\mathbf{X}|_{\mathcal{I}_{j-1}}\|_{2,1} \leq \frac{1}{\sqrt{R + K}} \|\mathbf{X}\|_{2,1}.$$

Also, we have $\lambda_0 = \|\mathbf{X}|_{\mathcal{I}_0}\|_F \leq \|\mathbf{X}\|_F$. Since $\mathbf{X} \in Q$, we conclude that

$$\sum_{j=0}^J \lambda_j \leq \left[\|\mathbf{X}\|_F + \frac{1}{\sqrt{R + K}} \|\mathbf{X}\|_{2,1} \right] \leq 1. \quad \blacksquare$$

PROOF OF LEMMA 2

Let $\mathbf{y}^{(i)}$ denote the i^{th} column of matrix \mathbf{Y} and $\mathbf{r}^{(i)}$ denote the i^{th} column of matrix \mathbf{R} , $i = 1, 2, \dots, L$. Then we have from Lemma 2 of [23]

$$\left(1 - \frac{\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}{1 - \delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}} \right) \|\mathbf{y}^{(i)}\|_2 \leq \|\mathbf{r}^{(i)}\|_2 \leq \|\mathbf{y}^{(i)}\|_2.$$

Summing the above relation, we obtain

$$\left(1 - \frac{\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}{1 - \delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}\right) \sum_{i=1}^L \|\mathbf{y}^{(i)}\|_2^2 \leq \sum_{i=1}^L \|\mathbf{r}^{(i)}\|_2^2 \leq \sum_{i=1}^L \|\mathbf{y}^{(i)}\|_2^2.$$

Equivalently, we have,

$$\left(1 - \frac{\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}{1 - \delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}\right) \|\mathbf{Y}\|_F \leq \|\mathbf{R}\|_F \leq \|\mathbf{Y}\|_F. \quad \blacksquare$$

REFERENCES

- [1] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse Solutions to Linear Inverse Problems with Multiple Measurement Vectors,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
- [2] I. F. Gorodnitsky, J. S. George, and B. D. Rao, “Neuromagnetic Source Imaging with FOCUSS : A Recursive Weighted Minimum Norm Algorithm,” *J. Electroencephalog. Clinical Neurophysiol.*, vol. 95, no. 4, pp. 231–251, Oct. 1995.
- [3] I. F. Gorodnitsky and B. D. Rao, “Sparse Signal Reconstruction from Limited Data using FOCUSS: A Re-weighted Minimum Norm Algorithm,” *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [4] D. Malioutov, M. Cetin, and A. S. Willsky, “A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
- [5] Petre Stoica and Randolph L. Moses, *Introduction to Spectral Analysis*, Upper Saddle River, N.J. Prentice Hall, 1997.
- [6] S. F. Cotter and B. D. Rao, “Sparse Channel Estimation via Matching Pursuit with Application to Equalization,” *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
- [7] Joel A. Tropp, “Algorithms for simultaneous sparse approximation. Part II: Convex relaxation,” *Signal Processing*, vol. 86, no. 3, pp. 589 – 602, 2006, Sparse Approximations in Signal and Image Processing Sparse Approximations in Signal and Image Processing.
- [8] D. P. Wipf and B. D. Rao, “An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, July 2007.
- [9] Zhilin Zhang and B. D. Rao, “Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sept 2011.
- [10] M. Elad and I. Yavneh, “A Plurality of Sparse Representations is Better Than the Sparsest One Alone,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4701 –4714, Oct. 2009.
- [11] S. K. Ambat, S. Chatterjee, and K.V.S. Hari, “Fusion of algorithms for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3699–3704, July 2013.
- [12] Sooraj K. Ambat, S. Chatterjee, and K.V.S. Hari, “Fusion of Algorithms for Compressed Sensing,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 5860–5864.
- [13] Sooraj K. Ambat, S. Chatterjee, and K.V.S. Hari, “A Committee Machine Approach for Compressed Sensing Signal Reconstruction,” *IEEE Trans. Signal Process.*, Accepted 2014.
- [14] Sooraj K. Ambat, Saikat Chatterjee, and K.V.S. Hari, “Progressive Fusion of Reconstruction Algorithms for Low Latency Applications in Compressed Sensing,” *Signal Processing*, vol. 97, no. 0, pp. 146 – 151, 2014.

- [15] Sooraj K. Ambat, Saikat Chatterjee, and K.V.S. Hari, “Fusion of Greedy Pursuits for Compressed Sensing Signal Reconstruction,” in *20th European Signal Processing Conference 2012 (EUSIPCO 2012)*, Bucharest, Romania, Aug. 2012.
- [16] D. Needell and J. A. Tropp, “CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples,” *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301 – 321, 2009.
- [17] Remi Gribonval, Holger Rauhut, Karin Schnass, and Pierre Vandergheynst, “Atoms of All Channels, Unite! Average Case Analysis of Multi-Channel Sparse Recovery using Greedy Algorithms,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 655–687, 2008.
- [18] E. van den Berg and M. Friedlander, “Probing the Pareto Frontier for Basis Pursuit Solutions,” *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.
- [19] E. van den Berg and M. P. Friedlander, “SPGL1: A Solver for Large-scale Sparse Reconstruction,” June 2007, <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [20] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), Circulation Electronic Pages: <http://circ.ahajournals.org/cgi/content/full/101/23/e215> PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.
- [21] Sooraj K. Ambat, Saikat Chatterjee, and K.V.S. Hari, “A Committee Machine Approach for Compressed Sensing Signal Reconstruction,” *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1705–1717, Apr. 2014.
- [22] Sooraj K. Ambat, Saikat Chatterjee, and K.V.S. Hari, “Progressive Fusion of Reconstruction Algorithms for Low Latency Applications in Compressed Sensing,” *Signal Processing*, vol. 97, no. 0, pp. 146 – 151, Apr. 2014.
- [23] Wei Dai and O. Milenkovic, “Subspace Pursuit for Compressive Sensing Signal Reconstruction,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230 –2249, May 2009.